Bolt Beranek and Newman Inc.

Report No. 5872

Quarterly Technical Report No. 1
October 16, 1983-January 15, 1984
Development of a Butterfly Multiprocessor Test Bed
Description of Butterfly Components

March 1985

Prepared for:
Defense Advanced Research Projects Agency
Engineering Applications Office

Quarterly Technical Report No. 1
October 16, 1983 - January 15, 1984

Development of a BUTTERFLY (TM) Multiprocessor Testbed

March 1985

Prepared for:

Dr. Clinton Kelly, Director
Defense Advanced Research Projects Agency
Engineering Applications Office
Arlington, Virginia 22209

Note: This research was sponsored by the Defense Advanced Research Projects Agency (DoD) under ARPA Order No. 4906, Contract No. MDA903-84-C-0033 issued by the Department of Army, Defense Supply Service: Washington, Washington, D.C. 20310.

Table of Contents

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 5872	2. GOVT ACCESSION NO. A153692 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Development of a Butterfly Multiprocessor Test Bed: Description of Butterfly Components	5. TYPE OF REPORT & PERIOD COVERED: Quarterly Technical Report Oct.16,1983-Jan.15,1984
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): J. Goodhue, E. Starr	8. CONTRACT OR GRANT NUMBER(s): MDA 903-84-C-0033
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Bolt Beranek and Newman, Inc. 10 Moulton St., Cambridge, MA 02238	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: ARPA Order No. 4906
11. CONTROLLING OFFICE NAME AND ADDRESS: DARPA 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: March, 1985
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiprocessor, Parallel Processor, Butterfly Switch, Packet Switching

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the hardware subassemblies used in the Butterfly Multiprocessor.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FIGURES

1. Introduction

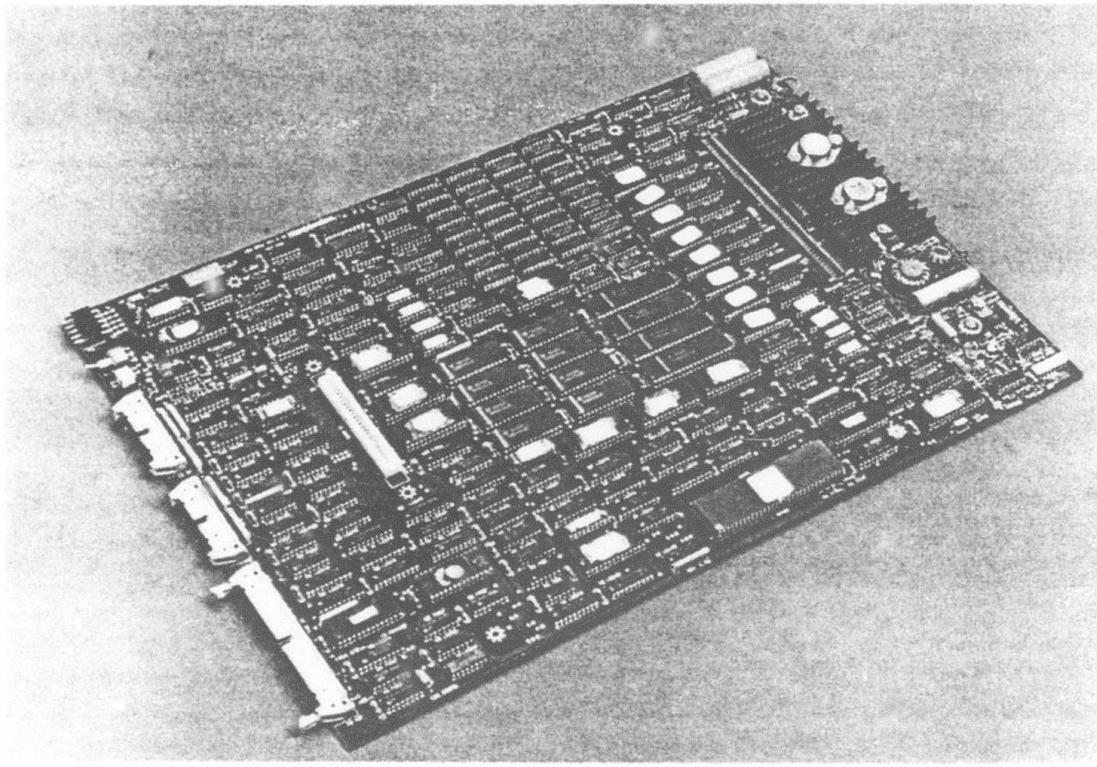
In this is the first in a series of Quarterly Technical Reports on the development of a 128-node Butterfly (TM) testbed under Contract No. MDA903-84-C-0033. In this Quarterly Report we present descriptions of the major components of the Butterfly Parallel Processor: the Processor Node (BPN), the MSI Switch Node (BSN), the VLSI Switch Node (BVSN), the Butterfly I/O Board (BI1), the Multibus Adapter (BMA), and the Butterfly Fantail (BFAN). This report updates and expands BBN Report No. 5284, which is an earlier summary of major Butterfly components. At the end of the report, there is a set of tables that summarize the important characteristics of each component. Keywords include: Parallel processor, Butterfly switch, and Packet switching.

2. Processor Node

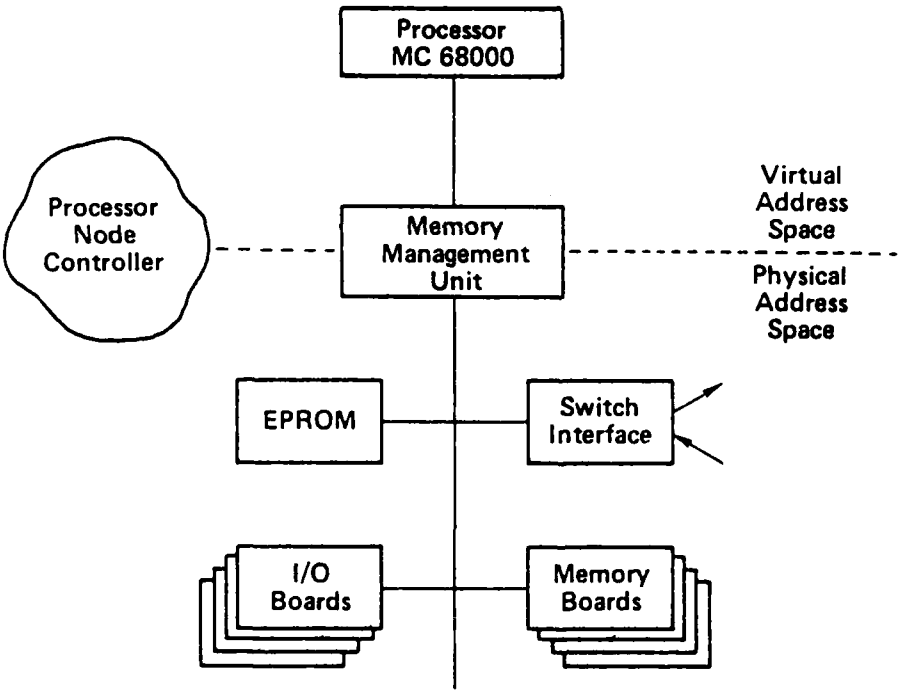
The Processor Node (BPN) is the source of computational and memory resources in the Butterfly (TM) Parallel Processor. Processor Nodes connected to the Butterfly Switch form the core of the machine. Figure 1 is a photograph of the 12 in. by 18 in. printed circuit board that implements the Butterfly Processor Node. A block diagram is shown in Figure 2. The Processor Node contains a Motorola MC68000 microprocessor, main memory, an AMD 2901-based co-processor called the Processor Node Controller (PNC), memory management hardware, an I/O bus, and an interface to the Butterfly Switch.

All application and operating system software runs on the Motorola MC68000 on the Processor Node. The MC68000 accesses system resources through the Memory Management Unit (MMU). When the MC68000 makes a memory reference, the Processor Node controller uses the memory management unit to determine whether the target of the reference resides in local or non-local memory. For non-local references, the PNC handles all of the interactions required to complete the transaction. Among other things, this means that the only difference to a user between a reference to local memory and a reference to memory on some other Processor Node is that the remote reference takes a little longer to complete. As a result, the memories on all of the Processor Nodes in a Butterfly system, taken together, appear to the application software as a single global memory.

The I/O bus on the Processor Node, commonly called the BIOlink, can be used to communicate directly with Butterfly I/O boards, or it can be used to connect to Multibus devices [IEEE83] through the Butterfly Multibus Adapter. The I/O bus is 16 bits wide, and capable of executing approximately one million data transfer operations per second. In addition to data transfer, the bus protocol gives I/O devices access to the Event mechanism supported by the Processor Node Controller.



Butterfly Processor Node (BPN)
Figure 1



Block Diagram of Butterfly Processor Node
Figure 2

Up to four I/O boards may be attached directly to a Processor Node. The Multibus adapter can be used to connect a Processor Node to as many Multibus boards as the IEEE specification will allow. Every Processor Node has its own independent I/O bus.

All of the memory in a Butterfly system is physically located on Processor Nodes. The amount of memory on a Processor Node is one megabyte when the memory sockets are populated with 256K dynamic RAM chips, or 256 Kbytes if 64K RAMs are used. By attaching a daughter board, it is possible to expand the memory to 4 megabytes; however, daughter boards require one additional card slot. In addition to its random access memory, the Processor Node has 8 Kbytes of EPROM for power-on diagnostics, a bootstrap loader, and a low-level debugger.

The switch interface on the Processor Node supports the transfer of messages to and from the Butterfly Switch. It is composed of two separate finite-state machines: one for output to the switch and one for input from the switch. Each finite-state machine connects to the switch through a separate connector at the edge of the board. The switch interface interacts with the rest of the Processor Node through a pair of dual-ported memories. When a message is to be sent out across the switch, the Processor Node Controller sets up a parameter block in the appropriate dual-ported memory and notifies the output finite-state machine. When a message comes in from the switch, the input finite-state machine deposits it in a dual-ported memory and notifies the Processor Node Controller.

The Processor Node Controller is a 2901-based microcoded co-processor for the MC68000. The PNC executes 8 million, 64-bit micro-instructions per second from a 4K word, read-only microstore. The PNC data paths match the 16-bit width of the MC68000 data bus. The vertical edge connector near the the Processor Node power supply allows the attachment of a writable control store for PNC

microcode development and debugging.

The PNC has several functions. First, it operates the various control wires that transfer data between components of the Processor Node. In this role, the PNC is involved in every memory reference made by the MC68000. It controls the flow of address information through the Memory Management Unit, watches for reference errors, and operates the memory system.

The second function of the PNC is to initiate all messages that go out to the switch, and to process all messages that arrive from the switch. For example, it is the PNC that makes remote memory references transparent to the MC68000. When the MC68000 requests a word of memory on another Processor Node, the PNC places the remote Processor Node number and memory address in a parameter block, and tells the switch output finite-state machine to send a read request. The MC68000 waits for the reply, but the PNC may service I/O interrupts or other micro-interrupts. When the message reaches the destination Processor Node, the remote PNC makes the memory reference and sends back a reply message. When the reply arrives, the value of the desired memory location is handed to the MC68000 as though it had been in local memory. Because the hardware is heavily pipelined, the roundtrip time for a remote memory reference is less than 4 microseconds. In addition to single word transfers, the PNC can be instructed to transfer blocks of memory between any pair of Processor Nodes in the machine. These transfers happen at the bandwidth of the switch (32Mbit sec).

Another function of the PNC is to augment the functionality of the MC68000 for parallel processing applications. This is done by programming the PNC to perform a variety of indivisible primitive operations. Examples of these primitives include a test-and-add instruction for maintaining reference counts and spin-locks, an Event mechanism, indivisible queueing operations,

and process scheduler primitives that work in concert with the queueing and event mechanisms to provide efficient communication and synchronization between application software modules. Use of these functions is described in the Chrysalis Operating System Manual. (LAR84)

The Processor Node has two LEDs on the edge of the board. One of these, next to the toggle switch that controls the on-board switching power supply, is green when the power supplies on the Processor Node and all attached I/O and memory boards are supplying the correct voltages. The second LED is red and is controlled by software. By convention, the red light is turned on at power up and whenever the Processor Node is reset; it is turned off when the Processor Node completes a built-in diagnostic test that indicates it is functioning properly.

3. MSI Switch Node

The Butterfly Switch combines techniques of packet-switching and sorting networks. The topology of the Butterfly Switch is similar to that of the bitonic sorting network described in BAT68. Its operation is much like that of a packet-switching network. It has been named the "Butterfly Switch" after the Fast Fourier Transform Butterfly which it resembles. The Butterfly Switch can be expanded to accommodate any number of processors.

The Butterfy Switch is implemented by an array of Switch Nodes connected to each other and to the Processor Nodes in the system. For reference, a block diagram of a 16-by-16 Butterfly Switch is shown in Figure 3. Logically, an individual Switch Node is a four-input four-output crossbar switch. In Figure 3, each input to a Switch Node on the left-hand side of the switch is connected to the output interface of a Processor Node. Similarly, each output from a Switch Node on the right-

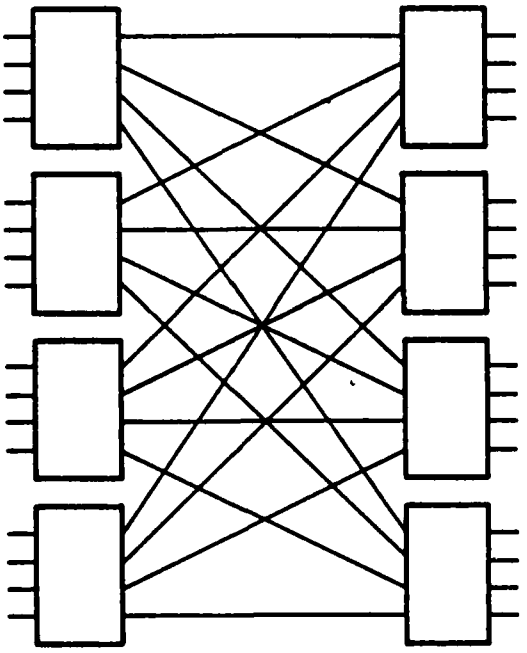
hand side of the switch is connected to the input of a Processor Node. Data flows through the switch from left to right.

For small Butterfly configurations, the MSI Switch Node (BSN) shown in Figure 4 is used as the basis for the switch. Each BSN is a self-contained unit with a switching power supply, connectors, drivers and receivers, and the logic that implements the switch algorithm. As shown in in the photograph, there are eight 26-pin connectors arranged in two columns at the edge of the board. The top four connectors (towards the left in the picture) are inputs to the Switch Node. The bottom four are outputs. The switch logic implements the routing, timing, flow control, and collision resolution mechanisms needed to route packets reliably through a single Switch Node.

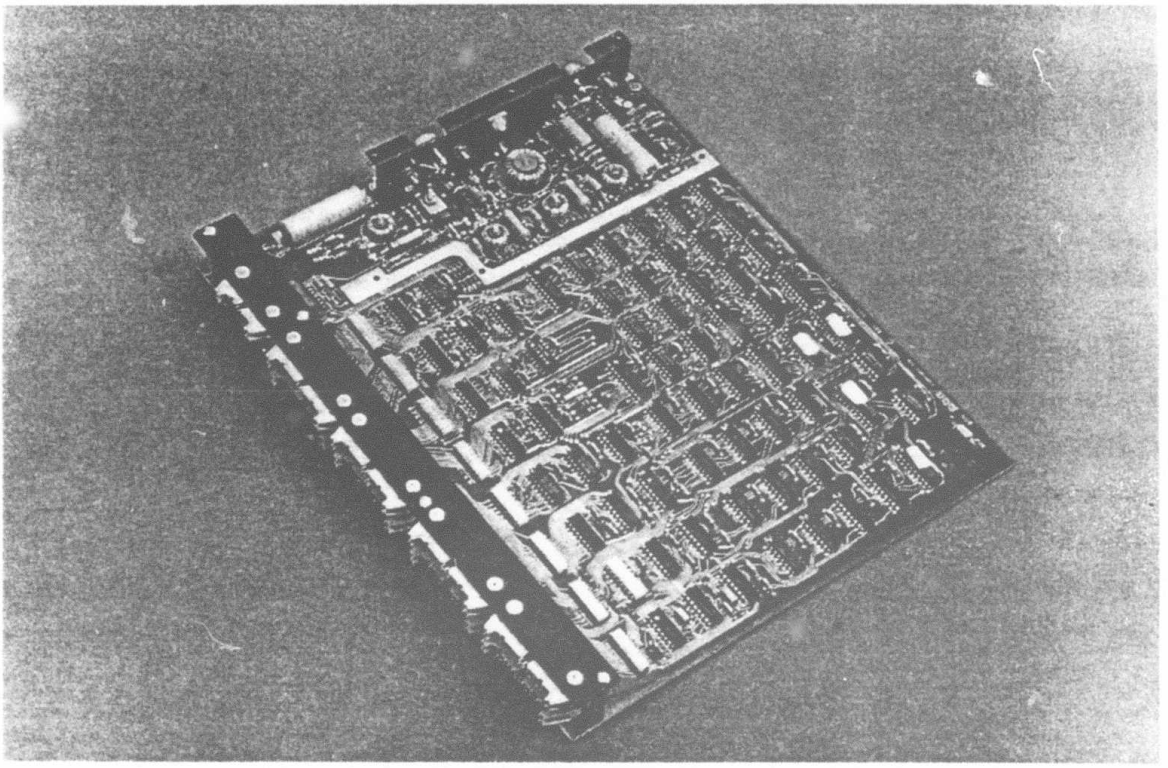
In addition to routing data, the Switch Node distributes a clock signal and a systemwide reset signal to the Processor Nodes. The Switch Node also uses the clock to regenerate signals passing through the switch. The clock and reset signals are supplied by a clock card (BCLK) to the Switch Node through the 16-pin connector near the top of the board.

The switching power supply on the BSN is similar to those used on other Butterfly boards. It is controlled by the toggle switch mounted on the front edge of the card. The LED next to the toggle switch is a "power-on" indicator.

In large machines, there may be significant distance between switch cards or between switch cards and Processor Nodes. In order to operate reliably at high speeds with immunity from ground reference problems, all of the signals on switch cables are driven differentially.



16-Port Butterfly Switch
Figure 3

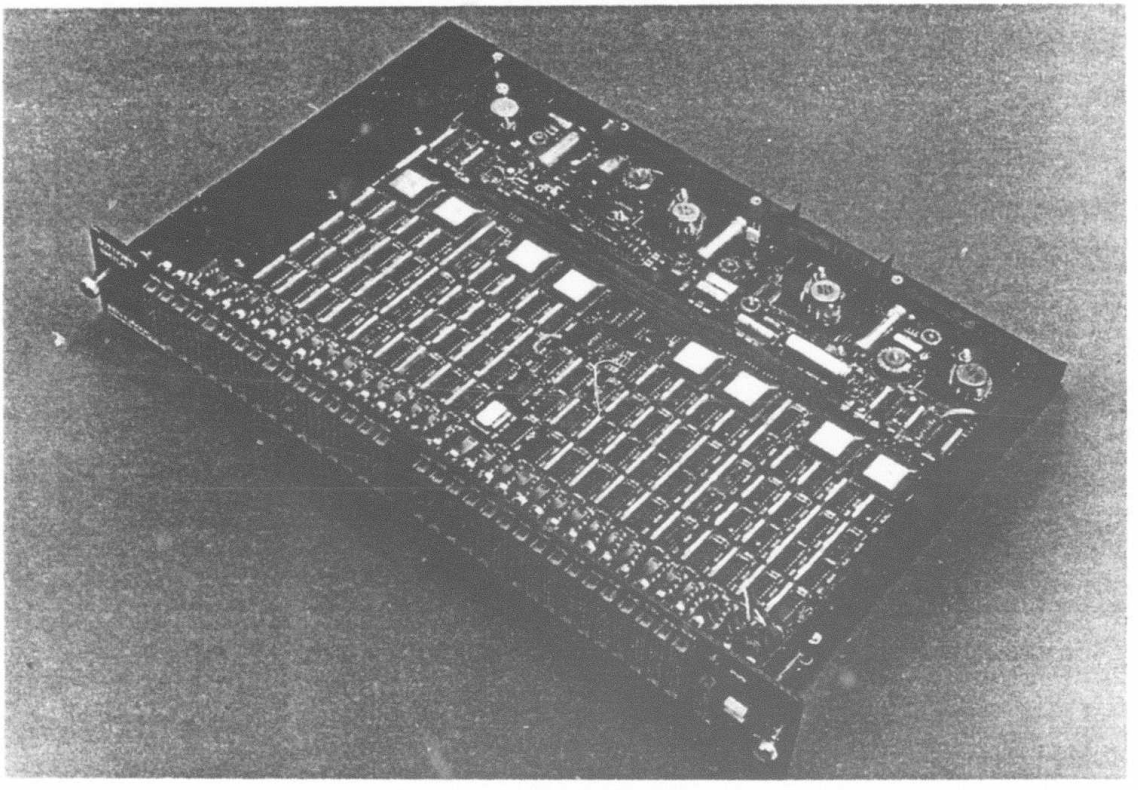


MSI Switch Node (BSN)
Figure 4

4. VLSI Switch Node

The Butterfly VLSI Switch Node (BVSN) is shown in Figure 5. It uses VLSI technology to implement a 16-by-16 Butterfly Switch on a single printed circuit board. Thus, it serves as a functional replacement for the eight MSI Switch Nodes (BSN) represented by the block diagram in Figure 3. Each of the custom VLSI chips near the rear of the board implements the 4-by-4 switching logic of the BSN. A single BVSN can be used to support a 16-processor Butterfly machine, or it can be cabled together with other BVSNs to form a switch for a larger machine.

The board is 12 in. by 18 in. (the same size as a Butterfly Processor Node). A 3 in. by 19 in. metal panel mounted at the front of the board provides mechanical support for 32 connectors (16 input and 16 output). The connectors mate to mass terminated cables which carry signals to Processor Nodes and other Switch Nodes. On-board clock and reset logic are included for systems that require only one BVSN. The reset logic is activated by the rocker switch on the front panel. In larger systems, the BVSN accepts external clock and reset signals through the 16-pin connector on the front panel.



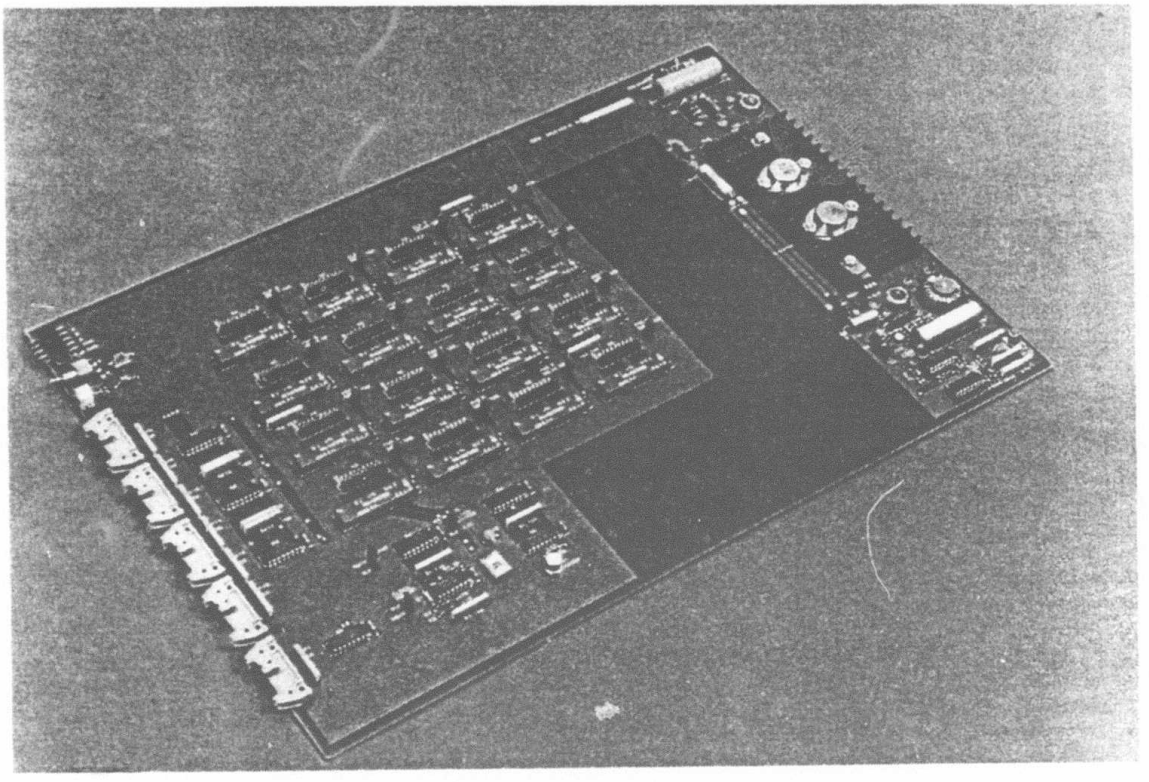
Butterfly VLSI Switch Node
Figure 5

5. Butterfly Clock

Since the Butterfly Parallel Processor is a fully synchronous machine, all Processor Nodes and Switch Nodes must use the same clock source. For systems based on a single VLSI switch card, the clock source is included on the switch card. For systems with multiple BVSN or BSN cards, the Butterfly Clock (BCLK) is used to distribute a clock signal from a central source. The printed circuit board that implements the Butterfly Clock is shown in Figure 6.

The Butterfly Clock operates in two modes: master mode, and slave mode. In master mode, it transmits a clock signal derived from an on-board oscillator through the four connectors nearest the top of the board in the photograph. A system reset signal, activated by the toggle switch near the top of the board, is distributed along with the clock signal. These connectors mate with mass-terminated cables that carry the the clock and reset signals around the machine. All of the BCLK outputs are source terminated. Each one is capable of driving two sets of receivers. Thus, a single BCLK is capable of supplying clock and reset signals to a total of eight switch cards. As noted earlier, the switch cards redistribute these signals to the Processor Nodes. All of the signals generated by the BCLK are differential ECL.

Butterfly systems with more than 64 processors use more than eight BVSNs. To accommodate systems of this size, the clock card operates in a slave mode. In slave mode, the BCLK takes in clock and reset signals through the fifth connector on the front edge of the board, regenerates them, and passes them through to its four output connectors. This makes it possible to distribute clock and reset signals through a tree of clock cards. The root node operates in master mode and sources the clock signal used by all of the other elements in the system. The remaining clock cards operate in slave mode, accepting signals from the previous level in the tree and passing them on to the next



Butterfly Clock Card
Figure 6

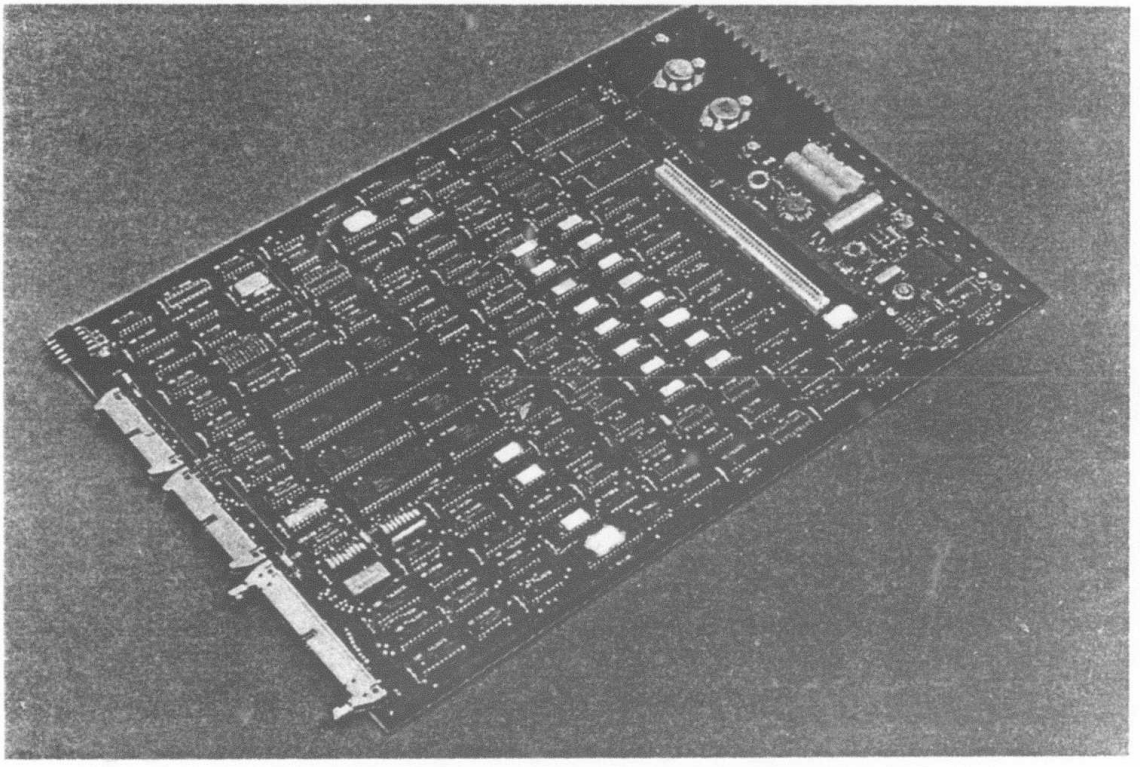
level down.

6. Butterfly I/O Board

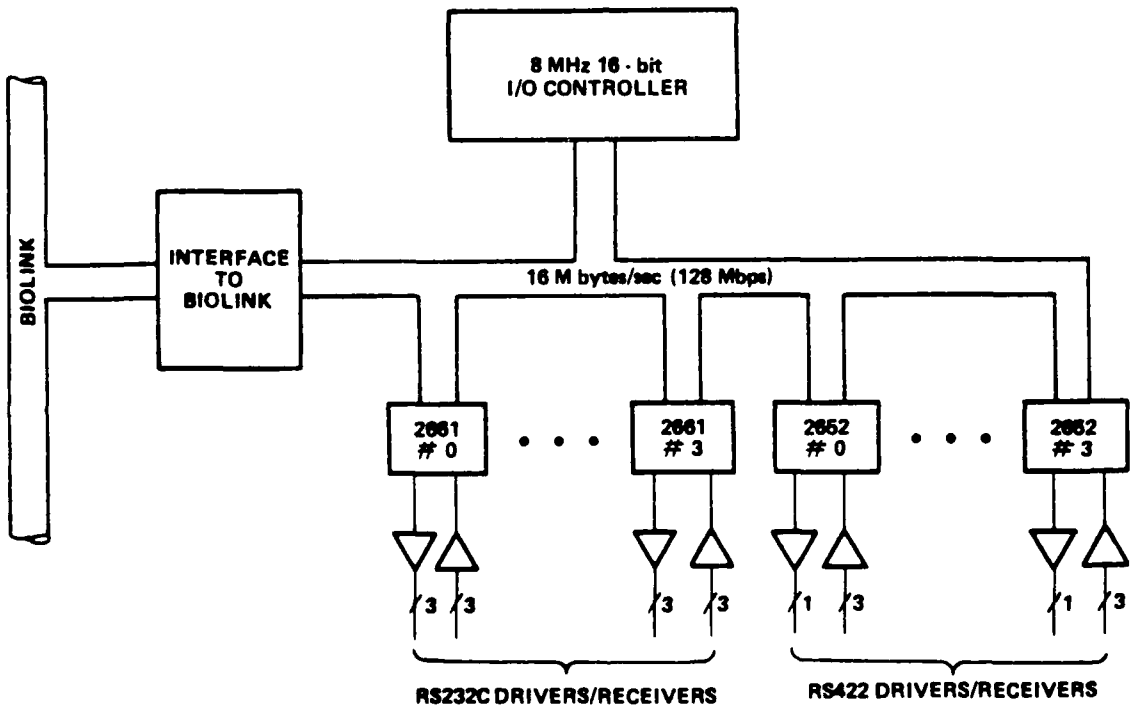
The Butterfly I/O Board (BI1) was developed to support a large number of synchronous and asynchronous I/O channels. Each Butterfly I/O Board includes four character-asynchronous I/O channels and four synchronous I/O channels. The asynchronous channels operate at speeds up to 38.4 kilobits per second, and are well suited to low-speed I/O devices such as terminals and local load devices. The synchronous channels operate at speeds up to 2 megabits per second and are used to service high-bandwidth devices. Figure 7 shows a photograph of the printed circuit board that implements the BI1, and Figure 8 is a simple block diagram.

The BI1 implements a sophisticated DMA mechanism for transferring data between the synchronous I/O channels and the main memory of the Processor Node. To support this mechanism, the BI1 uses a bit-slice microprocessor based on the AMD2901; about one third of the BI1 is devoted to this micromachine and associated hardware. The vertical edge connector near the rear of the board allows the attachment of a writable control store which is used for microcode development and debugging.

The second major section of the BI1 board is the hardware that implements the individual I/O channels. Associated with each asynchronous channel is a Signetics 2661 UART and a set of EIA RS-232 drivers and receivers. In addition to transmitting and receiving data, the asynchronous channels support various modem control signals. All of these signals are available through a 34-pin connector which mates to a mass-terminated ribbon cable. Associated with each synchronous



Butterfly I/O Board (BI1)
Figure 7



Block Diagram of Butterfly I/O Board
Figure 8

channel is a Signetics 2652 serial communications controller and a set of EIA RS-422 drivers and receivers. The Signetics 2652 implements much of the necessary framing required at the bit level for the HDLC protocol. Each synchronous channel supports transmit and receive data signals, plus transmit and receive clocks. All of these signals are available through a 40-pin connector which mates to another mass-terminated ribbon cable.

The remainder of the logic on the board implements the interface to the bus (BIOlink) that connects the BI1 to a Processor Node. The required signals come off the BI1 through a 50-pin connector which mates to a mass-terminated ribbon cable.

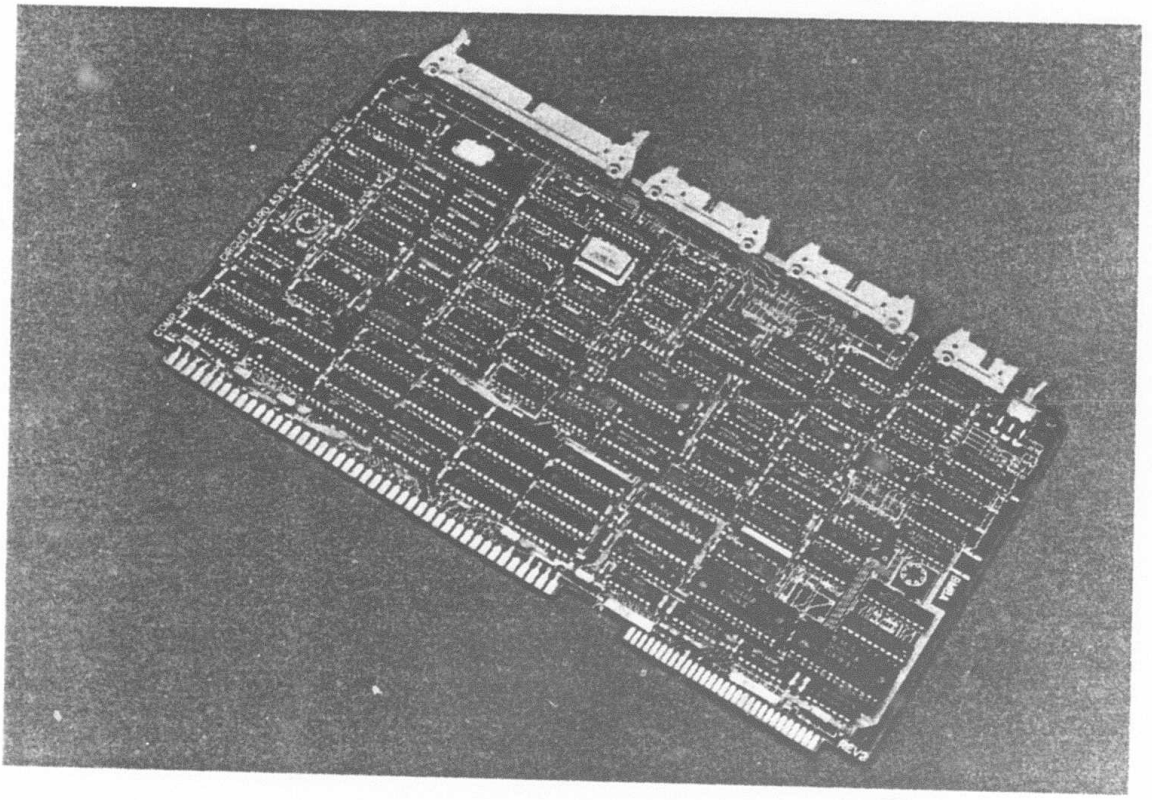
Like the Processor and Switch Nodes, the BI1 contains an on-board switching power supply, but there is no on-board power switch. Instead, the BI1 power supply is controlled by the Processor Node to which it is attached. There is no power indicator on the I/O board. If the I/O board power supply is not supplying the correct voltages, the power indicator light on the Processor Node to which it is attached will not come on. The power connector for the BI1 is located on the same edge of the board as the I/O connectors.

7. Multibus Adapter

The Butterfly Multibus Adapter (BMA) allows the connection of Butterfly Processor Nodes to I/O boards that conform to IEEE Standard 796 (Multibus). As shown in Figure 9, the BMA is a single four-layer printed circuit card that plugs into a Multibus backplane. With the exception of some of the "unused" signals on the Multibus P2 connector, the BMA conforms electrically, mechanically, and functionally to IEEE 796 specifications. The BMA connects to a Processor Node through the BIOlink in the same manner as the BI1.

The primary function of the BMA is to serve as a data channel between a Butterfly Processor Node and a Multibus. This channel allows devices on the BIOlink to access memory and I/O addresses on the Multibus, and allows Multibus devices access to memory locations on the Processor Node to which it is attached. Devices on the Multibus cannot access memory across the Butterfly Switch. Multibus devices can also use the Event mechanism supported by the Processor Node Controller. In addition to data transfer and event posting, the BMA supports Multibus interrupt processing. Each Multibus interrupt level can be mapped into one of two MC68000 interrupts in vectored or non-vectored format. Each Multibus interrupt level can be individually enabled. There is also a general interrupt disable function.

The BMA includes some extra facilities for small Butterfly systems: a pair of USARTs, the clock/reset connector, and the null switch interface. The USARTs on the BMA make it possible to connect to a console and a load device without incurring the cost of a separate I/O card. The null switch interface emulates a Butterfly switch. Hence, a single-processor system with a BMA does not require a BCLK or a switch card. For larger systems, the clock/reset connector on the BMA emulates the BCLK. Therefore, systems with up to four processors can operate without a BCLK, if



Butterfly Multibus Adapter
Figure 9

they include a BMA.

In addition to its primary functions, the BMA provides 2 Kbytes of CMOS RAM that can be attached to a battery in the Multibus chassis. It also contains a watchdog timer that is capable of asserting the reset signals on the null switch interface and the clock/reset connector. The reset can also be controlled through a pin on the P2 connector.

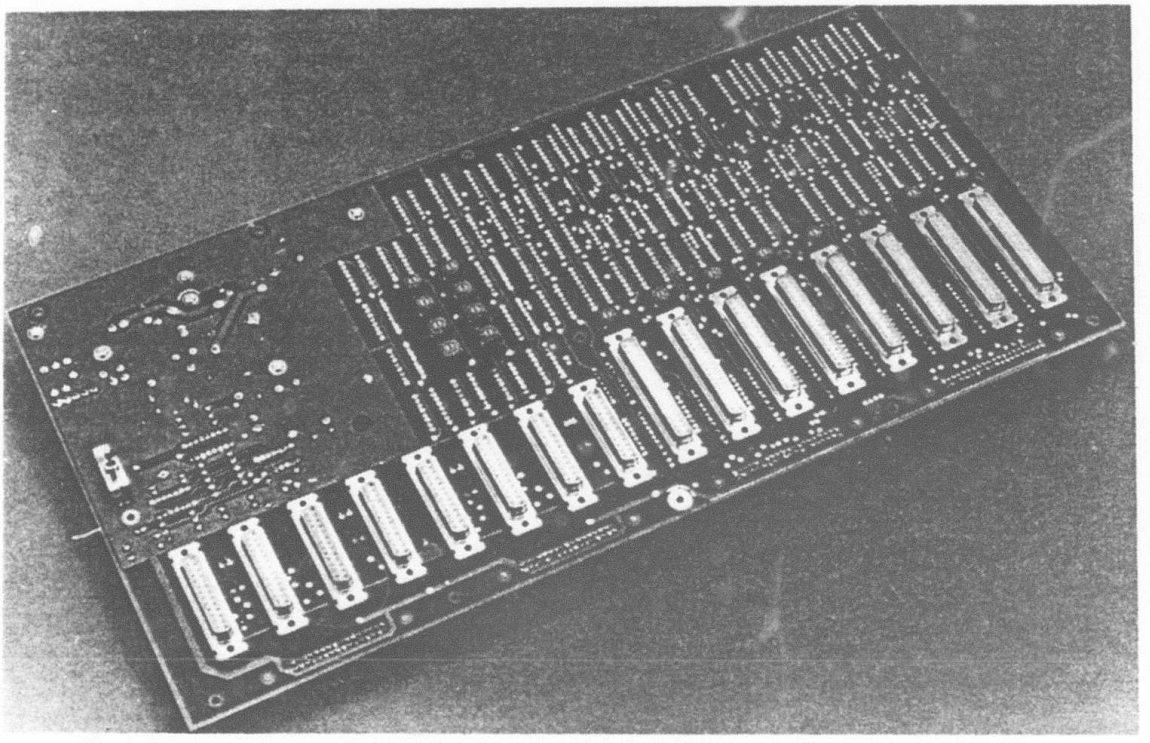
In Figure 9, the toggle switch at the top of the card controls the reset signals issued by the null switch and clock/reset connectors. The 16-pin connector is the clock/switch connector; the two connectors below it carry the signals for the null switch interface. The 68-pin connector at the bottom of the card mates to a BIOlink cable. The USART signals appear on the Multibus side of the card and are accessible through the P2 connector.

8. Butterfly Fantail

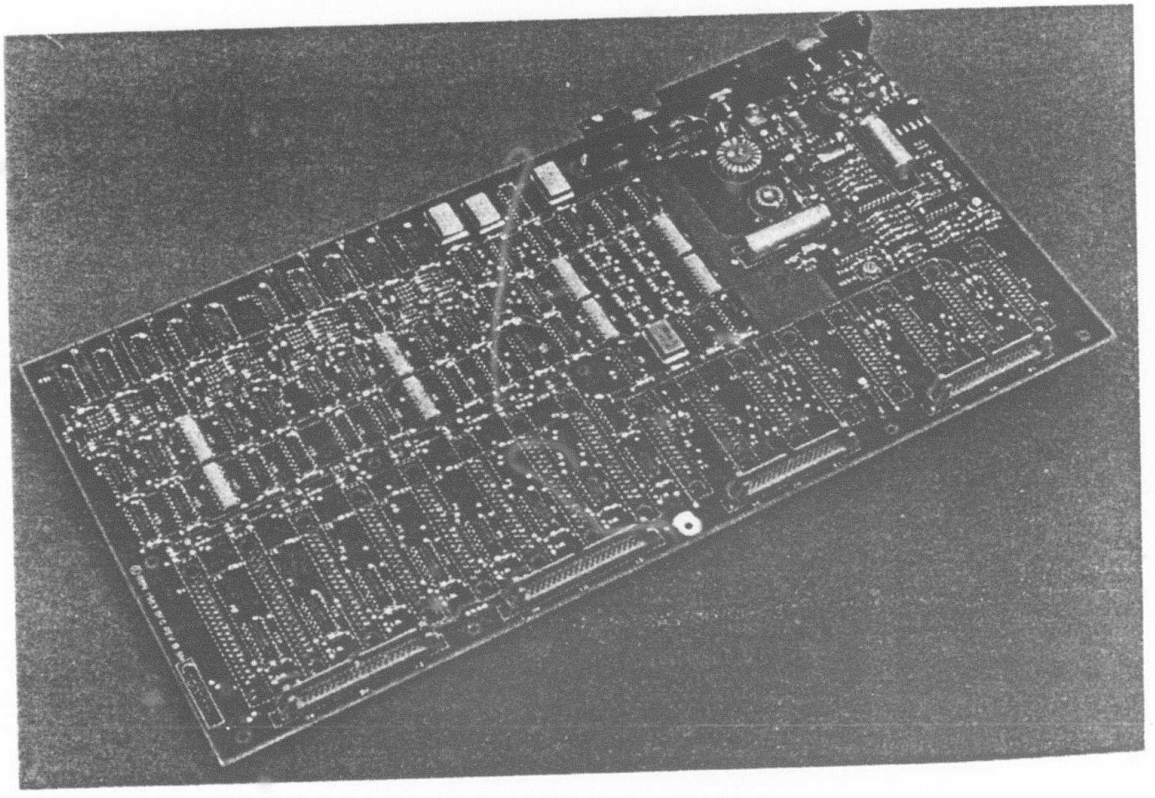
The Butterfly Fantail (BFAN) mates to the mass terminated cables that carry I/O data from the BI1. It routes the signals on these cables to separate 25 and 37-pin connectors. The BFAN is also capable of providing clock signals when they are needed for synchronous operation. A single Butterfly Fantail provides enough connectors for two BI1s. For each BI1 asynchronous channel, the BFAN provides a 25-pin connector with the pinout and form factor specified for a DCE by EIA standard RS-232. For each BI1 synchronous channel, the BFAN provdes a 37-pin connector with the pinout and form factor specified for a DCE by EIA standard RS-449. Also associated with each synchronous channel is a small rotary switch that selects among various clocking options.

The front of the BFAN is shown in Figure 10. The rear is shown in figure 11. On the front, there are eight 25-pin connectors, eight 37-pin connectors, a clock selection switch for each of the synchronous channels, and eight additional switches that control clock generators for the synchronous channels. The toggle switch on the lower right hand corner controls the power supply for the active circuitry on the BFAN.

A printed circuit board is attached to the rear of the sheet metal panel that covers the BFAN. This board holds all of the connectors, switches, and active circuitry on the BFAN. As shown in Figure 10, there are four connectors at the top of the board, one for each group of four synchronous and asynchronous channels. These connectors have the same pinout as the corresponding connectors on the BI1 board. An on-board switching power supply generates the voltages required by the active circuitry on the Fantail.



Butterfly Fantail (Front)
Figure 10



Butterfly Fantail (Rear)
Figure 11

REFERENCES

BAT68 Batcher, K.E. AFIPS Conference Proceedings, Volume 32, 1968 Spring Joint Computer Conference

IEEE83 "IEEE Standard Microcomputer System Bus" IEEE STD-796-1983 Institute of Electrical and Electronics Engineers Inc. 1983

LAR84 "Chrysalis Operating System Manual" Larus, J. et al. 1984

Appendix A. Tables

TABLE 1: SUMMARY OF CHARACTERISTICS OF BPN2	
Processor	Motorola MC68000 operating at 8MHz
Processor Node Controller	Bit Slice 2901 operating at 8MHz; 4K words of microcode, 64 bits wide
Main Memory per Card	1 Mbyte total with 256K RAMs; 256k bytes total with 64k RAMs
I/O	up to four I/O cards can be connected to the I/O bus; peak transfer rate is 16 MBits/sec
Board Size	12" x 18", 6 layers
Indicators	Power LED (green); program-controlled LED (red)
Controls	On/Off toggle switch
TABLE 2: SUMMARY OF BSN CHARACTERISTICS	
Function	Implements 4x4 Butterfly Switch Node
Bandwidth	32 Mbits/sec per switch path
Outputs	differential ECL
Board Size	12"x18", 4 layers
Indicators	"power on" LED (green)
Controls	On/Off toggle switch
TABLE 3: SUMMARY OF BVSN CHARACTERISTICS	
Function	Implements 16x16 Butterfly Switch Onboard Switching Power Supply, clock, and reset logic
Bandwidth	32 Mbits/sec per switch path
Board Size	12"x18", 6 layers
Outputs	16 input/16 output; differential ECL
Indicators	Power-on (green), Reset (red)
Controls	Reset (rocker switch)

TABLE 4: CHARACTERISTICS OF BCLK	
Frequency	8 MHz
Distribution	4 output connectors to drive up to 8 BVSN or BSN
Modes	Master or Slave
Board Size	12"x18"; 2 layers
Indicators	Power-on (green), Reset (toggle switch)
Controls	Reset (red)

TABLE 5: CHARACTERISTICS OF BI1	
Functions	I/O connections to BIOlink onboard switching power supply
I/O Support	4 asynch channels @ up to 38.4 Kb/sec each (EIA RS-232) 4 synch channels @ up to 2 Mb/sec each (EIA RS-422)
Board Size	12"x18"; 4 layers
Indicators	coupled through BPN to which BI1 is connected

TABLE 6: CHARACTERISTICS OF BMA	
Function	Connects BIOlink to Multibus Clock/Reset, Watchdog Timer
I/O Support	IEEE STD796 (MULTIBUS) 2 channels RS-232 Asynch
Board Size	Std Multibus, 4 layers